



Single nucleotide polymorphism discovery in bovine pituitary gland using RNA-Seq technology

Pareek, Chandra Shekhar; Smoczynski, Rafal; Kadarmideen, Haja; Dziuba, Piotr; Blaszczyk, Pawel; Sikora, Marcin; Walendzik, Paulina; Grzybowski, Tomasz; Pierzchala, Mariusz; Horbanczuk, Jaroslaw; Szostak, Agnieszka; Ogluszka, Magdalena; Zierrzchowski, Lech; Czarnik, Urszula; Fraser, Leyland; Sobiech, Przemyslaw; Wasowicz, Krzysztof; Gelfand, Brian; Feng, Yaping; Kumar, Dibyendu

Published in:
P L o S One

DOI:
[10.1371/journal.pone.0161370](https://doi.org/10.1371/journal.pone.0161370)

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Pareek, C. S., Smoczynski, R., Kadarmideen, H., Dziuba, P., Blaszczyk, P., Sikora, M., Walendzik, P., Grzybowski, T., Pierzchala, M., Horbanczuk, J., Szostak, A., Ogluszka, M., Zierrzchowski, L., Czarnik, U., Fraser, L., Sobiech, P., Wasowicz, K., Gelfand, B., Feng, Y., & Kumar, D. (2016). Single nucleotide polymorphism discovery in bovine pituitary gland using RNA-Seq technology. *P L o S One*, 11(9), [e0161370]. <https://doi.org/10.1371/journal.pone.0161370>

RESEARCH ARTICLE

Single Nucleotide Polymorphism Discovery in Bovine Pituitary Gland Using RNA-Seq Technology

Chandra Shekhar Pareek^{1*}, Rafał Smoczyński¹, Haja N. Kadarmideen², Piotr Dziuba¹, Paweł Błaszczyk¹, Marcin Sikora¹, Paulina Walendzik¹, Tomasz Grzybowski³, Mariusz Pierzchała⁴, Jarosław Horbańczuk⁴, Agnieszka Szostak⁴, Magdalena Ogluszka⁴, Lech Zwierzchowski⁴, Urszula Czarnik⁵, Leyland Fraser⁵, Przemysław Sobiech⁶, Krzysztof Wąsowicz⁶, Brian Gelfand⁷, Yaping Feng⁷, Dibyendu Kumar⁷



1 Division of Functional Genomics in Biological and Biomedical Research, Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University, Torun, Poland, **2** Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, **3** Ludwik Rydygier Collegium Medicum, Institute of Forensic Medicine, Department of Molecular and Forensic Genetics, The Nicolaus Copernicus University, Bydgoszcz, Poland, **4** Institute of Genetics and Animal Breeding of the Polish Academy of Sciences, Jastrzebiec, Poland, **5** Faculty of Animal Bio-engineering, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland, **6** Faculty of Veterinary Medicine, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland, **7** Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States of America

OPEN ACCESS

Citation: Pareek CS, Smoczyński R, Kadarmideen HN, Dziuba P, Błaszczyk P, Sikora M, et al. (2016) Single Nucleotide Polymorphism Discovery in Bovine Pituitary Gland Using RNA-Seq Technology. PLoS ONE 11(9): e0161370. doi:10.1371/journal.pone.0161370

Editor: Peter J. Hansen, University of Florida, UNITED STATES

Received: April 15, 2016

Accepted: August 4, 2016

Published: September 8, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Eighteen FastQ sequence dataset of bovine pituitary gland transcriptome submitted to NCBI-SRA database were available at http://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=312148.

Funding: This research paper was financially supported by National Science Centre, Krakow, Poland, Project No. 2012/05/B/NZ2/01629; entitled "Analiza transkryptomów genomu *Bos taurus* przy zastosowaniu technologii sekwencjonowania kolejnej generacji".

* pareekcs@umk.pl

Abstract

Examination of bovine pituitary gland transcriptome by strand-specific RNA-seq allows detection of putative single nucleotide polymorphisms (SNPs) within potential candidate genes (CGs) or QTLs regions as well as to understand the genomics variations that contribute to economic trait. Here we report a breed-specific model to successfully perform the detection of SNPs in the pituitary gland of young growing bulls representing Polish Holstein-Friesian (HF), Polish Red, and Hereford breeds at three developmental ages viz., six months, nine months, and twelve months. A total of 18 bovine pituitary gland polyA transcriptome libraries were prepared and sequenced using the Illumina NextSeq 500 platform. Sequenced FastQ databases of all 18 young bulls were submitted to NCBI-SRA database with NCBI-SRA accession numbers SRS1296732. For the investigated young bulls, a total of 113,882,3098 raw paired-end reads with a length of 156 bases were obtained, resulting in an approximately 63 million paired-end reads per library. Breed-wise, a total of 515.38, 215.39, and 408.04 million paired-end reads were obtained for Polish HF, Polish Red, and Hereford breeds, respectively. Burrows-Wheeler Aligner (BWA) read alignments showed 93.04%, 94.39%, and 83.46% of the mapped sequencing reads were properly paired to the Polish HF, Polish Red, and Hereford breeds, respectively. Constructed breed-specific SNP-db of three cattle breeds yielded at 13,775,885 SNPs. On an average 765,326 breed-specific SNPs per young bull were identified. Using two stringent filtering parameters, i.e., a minimum 10 SNP reads per base with an accuracy $\geq 90\%$ and a minimum 10 SNP reads per base with an accuracy = 100%, SNP-db records were trimmed to construct a highly reliable SNP-db. This resulted in a reduction of 95,7% and 96,4% cut-off mark of constructed

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: SNP, Single nucleotide polymorphism; RNA-seq, RNA sequencing; HF, Holstein-Friesian; KASP™, kompetitive allele-specific PCR; BWA, Burrows-Wheeler Aligner; BCM-HGSC, Baylor College of Medicine Human Genome Sequencing Centre; MAS, marker assisted selection; GAS, gene assisted selection; GS, genomics selection; QTL, Quantitative trait locus; BTA, Bos Taurus autosome; indel, insertions and deletions; SNP-db, SNP database; HWE, Hardy Weinberg equilibrium; NGS, next generation sequencing; RIN, RNA Integrity Number; QC, quality control; TMEM9, transmembrane protein 9; DDX56, DEAD (Asp-Glu-Ala-Asp) box polypeptide 56; RNF167, ring finger protein 167; CGs, candidate genes; GHSR, growth hormone secretagogue receptor; IGF2BP2, insulin-like growth factor 2 binding protein 2; GAD1, Glutamate decarboxylase 1; IGFBP2, insulin-like growth factor binding protein 2; IGFBP5, insulin-like growth factor binding protein 5; MSTN, myostatin; LEPR, leptin receptor; APOA2, apolipoprotein A2; SLC44A5, solute carrier family 44, member 5; NTNG1, netrin G1; GHRHR, growth hormone releasing hormone receptor; IGF2BP3, insulin-like growth factor 2 binding protein 3; IGFBP1, insulin-like growth factor binding protein 1; IGFBP3, insulin-like growth factor binding protein 3; LEP, leptin; NPY, neuropeptide Y; NSIG1, insulin Induced gene 1; DPP6, dipeptidyl aminopeptidase-like protein 6; HGF, hepatocyte growth factor; IGFBP6, insulin-like growth factor binding protein 6; IGF-I, insulin-like growth factor 1; WNT10B, wingless-type MMTV integration site family member 10B; MYF6, myogenic factor 6; MYF5, myogenic factor 5; IGFBP7, insulin-like growth factor binding protein 7; CCSE1, coiled-coil serine-rich protein 1; PKD2, polycystic kidney disease 2; NCAPG, non-SMC condensin I complex, subunit G; KCNIP4, kv channel interacting protein 4; CAST, calpastatin; NPM1, nucleophosmin; PROP1, PROP paired-like homeobox 1; IGFBPL1, insulin-like growth factor binding protein-like 1; ATP6V1B2, ATPase, H⁺ transporting, lysosomal 9kDa, V1 subunit B2; IGF2R, insulin-like growth factor 2 receptor; RFX6, regulatory factor X, 6; MAP3K5, mitogen-activated protein kinase 5; GTF3C5, general transcription factor IIIC, polypeptide 5; RALGDS, ras association; SDC1, syndecan 1; DNMT3B, DNA (Cytosine-5-)methyltransferase 3 beta; GHRH, growth hormone releasing hormone; FOXA2, forkhead box A2; HNF4A, hepatocyte nuclear factor 4, alpha; PLAG1, pleiomorphic adenoma gene 1; NUCB2, nucleobindin 2; BTG4, B-cell translocation gene 4; MGC134087, Bos taurus UPF0686 protein C11orf1 homolog-like (LOC538766); ATM, ATM Serine/Threonine Kinase; UCP3, uncoupling protein 3; IGFBP1, immunoglobulin-

raw SNP-db. Finally, SNP discoveries using RNA-Seq data were validated by KASP™ SNP genotyping assay. The comprehensive QTLs/CGs analysis of 76 QTLs/CGs with RNA-seq data identified *KCNIP4*, *CCSE1*, *DPP6*, *MAP3K5* and *GHR* CGs with highest SNPs hit loci in all three breeds and developmental ages. However, *CAST* CG with more than 100 SNPs hits were observed only in Polish HF and Hereford breeds. These findings are important for identification and construction of novel tissue specific SNP-db and breed specific SNP-db dataset by screening of putative SNPs according to QTL db and candidate genes for bovine growth and reproduction traits, one can develop genomic selection strategies for growth and reproductive traits.

Introduction

In recent years, the development of RNA-seq technology has completely revolutionized the way of thinking in molecular biology, and allowed a more detailed insight into gene expression by examining expression level, sequence variation, gene structure, and strand-specificity simultaneously [1–4]. The bovine genome is about 3 GB (3 billion base pairs), and contains approximately 22,000 genes, of which 14,000 are common to all mammalian species. The genome assembly of a Hereford cow produced by the Baylor College of Medicine Human Genome Sequencing Centre (BCM-HGSC) has now allowed years of work in cattle QTL mapping to be associated with genes and other genomic features [5–6]. This high throughput genome sequencing and identification of abundant amount of DNA variants for the bovine genome have significantly facilitated the analysis of genetic variation among cattle breeds. However, very little is known regarding the transcriptomic expression of genes in cattle and the variation between breeds. However, very little is known regarding the SNP-driven transcriptomic expression of genes in cattle and the tissue level database of SNPs (SNP-db) that associate with expressed genes in the tissue. Therefore, importance of understanding single nucleotide variation in the mapped transcriptome is important to identify the trait-associated genomic mutations in shaping the phenotypes [7]. In this study, the pituitary gland transcriptomes of three cattle breeds: Polish HF, Polish Red, and Hereford were investigated. Polish HF and Polish Red are the native cattle breeds utilized both for dairy and beef productions, whereas the Hereford breed is used solely for beef production. It is well known that the pituitary gland transcriptome, due to its potential endocrinal role, has the major influence towards genetic improvement of economic traits, particularly the growth and reproduction trait. In general, the bovine pituitary gland is responsible for growth control at different stages of development. This involves complex interactions of several hormones and growth factors, acting in both an endocrine, and a paracrine or autocrine manner. In this context bovine growth and development in young animal ages, candidate genes such as the *pituitary growth hormone (GH or somatotropin)* and its receptor (*GHR*), *insulin-like growth factors (IGF)*, *IGF binding proteins (IGFBP)* and the *IGF-I receptor (IGF-IR)*, *myogenic factors (MYFs)*, and *somatostatin*, plays an essential role in postnatal growth regulation, especially in nutrient utilization [8–11]. Generally, in cattle breeding practice and selection programs, the maximum genetic response is favorable for traits of economic importance, which includes growth traits.

In animal genomics, RNA-Seq could also provide a powerful tool for high-resolution genomic analysis of tissues and cell populations of domestic animals to identify novel mutations and transcripts affecting the major economic traits [12]. Gene expression thus is a useful marker for economic traits, suggesting that RNA-Seq could play a role in domestic animal

like and fibronectin type III domain containing 1; CAPN2, calpain 2; PRDM16, PR domain containing 16; ZBED6, zinc finger, BED-type containing 6; LHX4, LIM homeobox 4; IGFLR1, IGF-like family member 1-like receptor; IGFL1, IGF-like family member 1-like; IGFBP4, insulin-like growth factor binding protein 4; IGFBP1, insulin-like growth factor 2 binding protein 1; FASN, fatty acid synthase; SREBF1, sterol regulatory element binding transcription factor 1; CARTPT, CART prepropeptide; GHR, growth hormone receptor; PRLR, prolactin receptor; IGF1R, insulin-like growth factor 1 receptor; GHRL, ghrelin/obestatin prepropeptide; MST1R, macrophage stimulating 1 receptor; DLK2, delta-like 2 homolog; MC4R, melanocortin 4 receptor; NPC1, niemann-pick disease, type C1; XYLT1, xylosyltransferase I; SH2B2, SH2B adaptor protein 2; SIRT1, sirtuin 1; IGF2, insulin-like growth factor 2; CAPN1, calpain 1; VEGFB, vascular endothelial growth factor B.

breeding and selection programs *viz.*, marker assisted selection (MAS), gene assisted selection (GAS) and genomics selection (GS) [13–15]. In animal genomics, detection of potential putative SNPs for economic traits has great potential in genetic improvement of all domestic live-stock animals, including cattle. In the past decade, implementations of MAS, GAS and GS were highly recommended to the global cattle breeding program worldwide. In last 5–6 years, several studies have reported on the identification of novel SNP discoveries using RNA-seq technologies in cattle [16–25] (Table 1). To best of our knowledge, no studies have reported about the detections of SNPs in bovine pituitary gland except, some of our preliminary results presented at an international conference [26]. A brief summary of the global RNA-seq studies on the detection of SNPs in cattle is illustrated in Table 1. In this study, we used mRNA-Seq to characterize and compare the bovine pituitary transcriptomes of the Hereford, Polish Red, and Polish HF breeds, with respect to single nucleotide variations, CGs, QTLs regions and to identify the novel breed-specific SNPs.

Results

mRNA sequencing and read alignment

To obtain the bovine pituitary gland transcriptome at a single-nucleotide resolution, two biological replicates of poly(A)-enriched mRNA of young bulls aged 6, 9, and 12 months from three cattle breeds were converted to barcoded strand-specific dUTP RNA-seq libraries, and sequenced on the Illumina NextSeq 500 sequencer. Sequencing generated a total of 113,882,3098 raw paired-end reads with a length of 156 bases. The reads were de-multiplexed to assign reads to each sequenced sample according to its index. An average of 63 million paired-end reads was obtained per library ($n = 18$). The FastQ sequence dataset of each library was submitted to NCBI-SRA experiment number SRS1296732 (http://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=312148). Using the BWA programme, breed-specific pituitary gland transcripts were mapped to the bovine reference genome (UMD3.1 assembly plus Y chromosome) and the results were summarized in Table 2 (Polish HF), Table 3 (Polish Red) and Table 4 (Hereford), respectively. Based on the BWA mapping programme under default condition, 93.04%, 94.39% and 83.46% of the mapped sequencing reads were properly paired for the Polish HF, Polish Red, and Hereford breeds, respectively (Tables 2–4).

SNP discoveries in cattle breeds

Using a stringent filtering parameter of read count (SAMtools package), with a minimum depth of 5, detection of putative SNPs were performed in three breeds. The Venn diagram is displayed to report the number of SNPs shared among three cattle breeds (Fig 1). For the detection of breed-specific putative SNPs, we utilized two stringent parameters, *viz.*, a minimum 10 SNP reads per base with an accuracy $\geq 90\%$, and a minimum 10 SNP reads per base with an accuracy = 100%. The SAMtools identified single base substitutions (SNPs) as well as small insertions and deletions (*indel*); however, only SNPs results were considered to publish in this paper. A total of 13.7 million breed-specific SNPs positions were detected with the RNA-Seq reads, with an average of 765,326 (~ 0.76 million) SNPs per young bull (S1–S18 Tables). This high percentage of novel SNPs suggests that a large fraction of the genetic variability present in cattle breeds still remains to be discovered. Furthermore, based on the S1–S18 Tables, the raw SNP database (SNP-db) of pituitary gland transcriptome representing the investigated young bulls of the Polish HF, Polish Red, and Hereford breeds was constructed (S19 Table). Using two stringent filtering parameters, SNP-db records were trimmed to construct a highly reliable SNP-db and resulting in reduction of 95,7% and 96,4% cut-off mark of constructed raw SNP-db (S19 Table).

Table 1. Review of literature on global tissue-specific bovine RNA-seq experiments based on SNP discoveries.

Transcriptome	Cattle breed	Economic trait	SNP discoveries	References
Milk	Holstein cows	Milk production trait	The SNP detection analysis revealed 100,734 SNPs in Holstein samples, and a large number of those corresponded to differences between the Holstein breed, and the Hereford bovine genome assembly Btau4.0. The number of polymorphic SNPs within Holstein cows was 33,045. The accuracy of RNA-Seq SNP discovery was tested by comparing SNPs detected in a set of 42 candidate genes expressed in milk that had been re-sequenced earlier using Sanger sequencing technology. Seventy of 86 SNPs were detected using both RNA-Seq and Sanger sequencing technologies.	[16]
Hypothalamus of a prepubertal and a postpubertal heifer.	Brangus (3/8 Brahman × 5/8 Angus)	Growth and fertility traits	Combined of BovineSNP50 BeadChip and RNA-seq data analysis identified SNPs from the hypothalamus of a prepubertal and a postpubertal Brangus heifer.	[17]
Leukocytes	Holstein, Jersey and Cholistani	Breed-specific RNA-seq experiment (None of the bovine economic traits were investigated)	Study identified breed-specific base changes in protein coding regions. Among 7,793,425 coding bases, only 165 differed between Holstein and Jersey, and 3,383 (0.04%) differed between Holstein and Cholistani, 817 (25%) of which resulted in amino acid changes in 627 genes.	[18]
Longissimus Muscle	Qinchuan beef cattle	Beef production	Study discovered that 30,618–31,334 putative single nucleotide polymorphisms were located in coding regions.	[19]
blastocysts	Bovine embryo	Fertility and reproduction trait	Study generated approximately 38 million single reads of 40 bp per embryo, with an average of approximately 29 million reads per sample(76.9%) mapped to annotated portions of the Ensembl-UMD3.1 genome. Expressed biallelic SNP variants with allelic imbalances were observed in 473 SNP, where one allele represented between 65–95% of a variant's transcripts.	[20]
Horn Cancer (HC) and Horn Normal (HN)	Kankrej	Complex disease trait (cancer)	Study found 9532 and 7065 SNPs as well as 1771 and 1172 Indels in HC and HN, respectively, from which, 7889 SNPs and 1736 Indels were uniquely present in HC, 5886 SNPs and 1146 Indels were uniquely present in HN and reported first time in Bos indicus.	[21]
Longissimus thoraci	Limousine bulls	Meat quality trait	For the meat quality trait 34,376 different SNPs were detected. Fifty-five percent of the SNPs were found in coding regions, and ~22% resulted in an amino acid change. Applying a very stringent SNP quality threshold, we detected 8,407 different high-confidence SNPs, 18% of which are non-synonymous coding SNPs.	[22]
Hypothalamus, pituitary gland, ovary, uterus, endometrium, longissimus dorsi muscle, adipose, liver	Brangus heifers, Brahman (Bos indicus) × Angus (Bos Taurus)	Fertility traits: age of first observed corpus luteum (ACL), first service conception (FSC), and heifer pregnancy (HPG)	Study revealed 25 QTL loci containing SNPs associated with fertility trait.	[23]

(Continued)

Table 1. (Continued)

Transcriptome	Cattle breed	Economic trait	SNP discoveries	References
Subcutaneous adipose tissue (fat) of a fetal, adult bulls, adult heifers and adult steers	Qinchuan	Beef production	Study detected 56,564 (fetal), 65,154 (adult bull), 78,061 (adult heifer), and 86,965 (adult steer) putative SNPs located in coding regions of the four pooled libraries of adipose tissue.	[24]
<i>Mycobacterium avium</i> subspecies <i>paratuberculosis</i> (MAP)	German Holstein cattle	Health disease trait	The genetic variability within a pool of seven genes (<i>LAMB1</i> , <i>DLD</i> , <i>WNT2</i> , <i>PRDM1</i> , <i>SOCS5</i> , <i>PTGER4</i> , and <i>IL10</i>) were investigate in a population of 324 German Holstein cattle (162 cases MAP positive and 162 controls MAP negative) by GWAS/ RNA-Seq analysis. Study discovered novel SNPs putatively associated with paratuberculosis susceptibility.	[25]

doi:10.1371/journal.pone.0161370.t001

Table 2. Transcriptome mapping of pituitary gland in Polish HF breed to bovine reference UMD3.1 genome assembly.

Age	Total	Mapped	Paired in sequencing	Read1	Read2	Properly paired	With itself and mate mapped	Singletons	Mate mapped to a different chromosome
	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶
6m	218,45	218,45	109,26	109,19	173,99	202,27	4,34	8,13	1,89
9m	159,83	159,83	79,84	79,98	132,45	148,29	2,38	3,17	0,59
12m	137,10	137,10	68,58	68,52	118,33	128,94	1,77	2,07	0,48
Total	515,38	515,38	257,38	257,70	424,78	479,51	8,49	13,38	2,97

doi:10.1371/journal.pone.0161370.t002

Breed-specific trimmed/filtered SNP-db showed the highest SNP counts in the Polish HF breeds with an average (per young bull) 880,442 raw SNPs, 39,730 SNPs (≥ 10 reads $\geq 90\%$) and 31,978 SNPs (≥ 10 reads = 100%), followed by 754,719 raw SNPs, 34,658 SNPs (≥ 10 reads $\geq 90\%$), 30,052 SNPs ≥ 10 reads = 100%) in Hereford breed, and 660,819 raw SNPs, 25,414

Table 3. Transcriptome mapping of pituitary gland in Polish Red breed to bovine reference UMD3.1 genome assembly.

Age	Total	Mapped	Paired in sequencing	Read1	Read2	Properly paired	With itself and mate mapped	Singletons	Mate mapped to a different chromosome
	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶
6m	108,49	108,49	54,27	54,22	94,08	102,68	1,36	1,14	628,78
9m	56,89	56,89	28,36	28,53	48,48	54,14	0,70	0,69	329,58
12m	50,00	50,00	25,03	24,96	42,02	46,48	0,71	0,78	287,20
Total	215,39	215,39	107,66	107,72	184,60	203,31	2,78	2,62	1 245,58

doi:10.1371/journal.pone.0161370.t003

Table 4. Transcriptome mapping of pituitary gland in Hereford breed to bovine reference UMD3.1 genome assembly.

Age	Total	Mapped	Paired in sequencing	Read1	Read2	Properly paired	With itself and mate mapped	Singletons	Mate mapped to a different chromosome
	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶	1x10 ⁶
6m	207,07	202,72	207,07	102,64	104,43	174,57	200,69	2,03	6,30
9m	105,39	103,66	105,39	52,01	53,37	82,16	102,80	0,86	6,31
12m	95,58	94,07	95,59	47,41	48,17	78,77	93,30	0,78	2,94
Total	408,04	401,97	408,04	202,06	205,98	335,51	396,79	3,67	15,54

doi:10.1371/journal.pone.0161370.t004

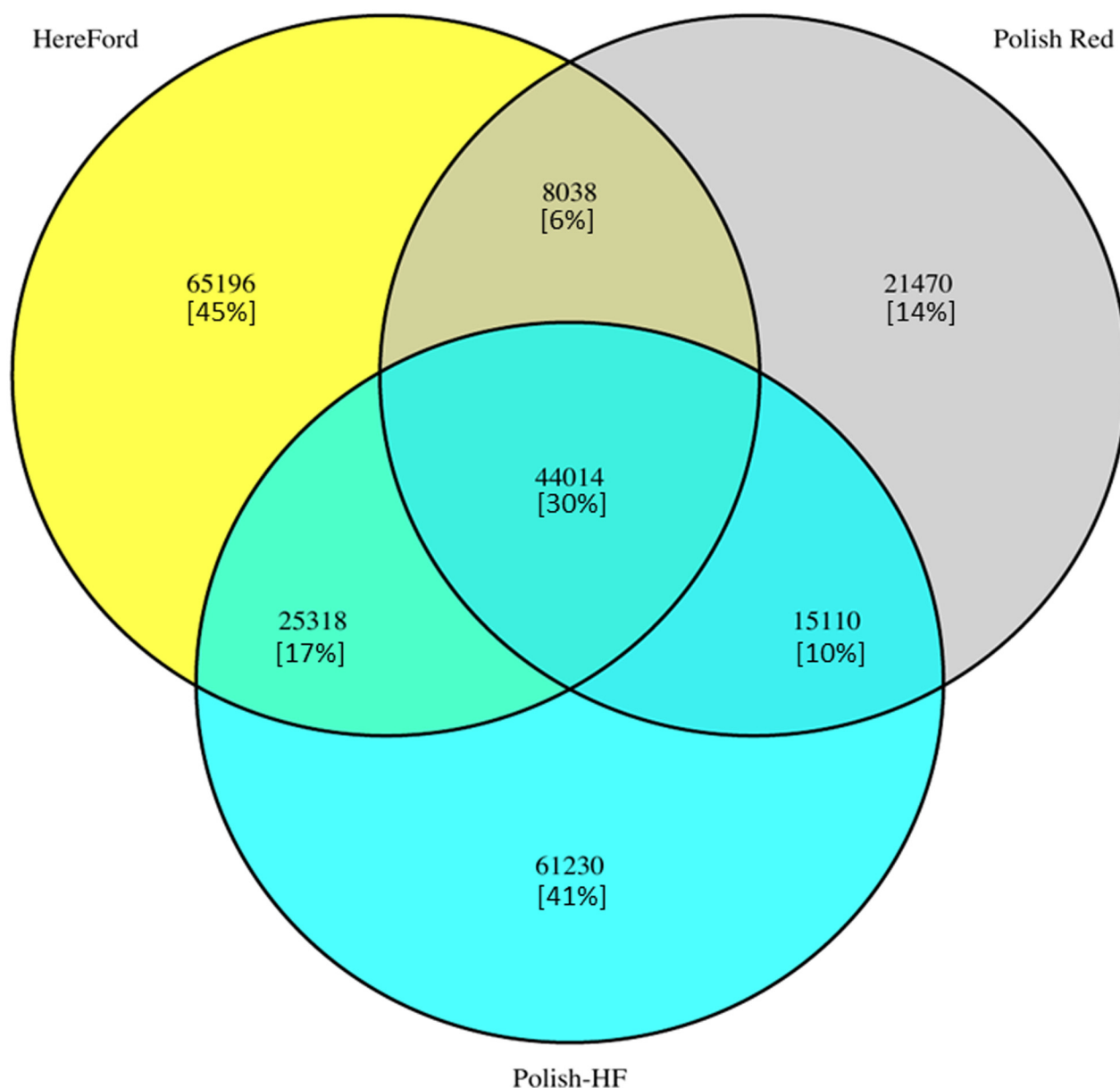


Fig 1. Venn diagram showing the number of SNP segregating within three cattle breeds.

doi:10.1371/journal.pone.0161370.g001

SNPs (≥ 10 reads $\geq 90\%$), 20,573 SNPs (≥ 10 reads = 100%) for Polish Red breed, respectively. Furthermore, the SNP-db showed the highest percentage of ≥ 10 reads $\geq 90\%$ and ≥ 10 reads = 100% filtered SNPs in Hereford breed (4.6% and 4.0%), followed by Polish HF breed (4.5% and 3.6%), and Polish Red breed (3.8% and 3.1%), respectively.

Based on their types and occurrences, filtered SNP-dbs [S20 and S21 Tables] were further categorized into the following four sub-categories, as shown in S22 Table.

1. SNPs represented in any of six samples per given breed (SNP-type designated as ANY)
2. SNP detected in all six samples per breed (SNP-type designated as Breed-specific).

3. SNP detected in any of six samples per given breed, not representing any other breed (SNP-type designated as ANY UNIQUE)
4. SNP detected in all six samples per breed, not representing any other breed (SNP-type designated as Breed-specific UNIQUE).

The [S22 Table](#) showed that SNP categorization substantially reduced the SNP-db, resulting in highly reliable breed-specific, and breed-specific unique SNP-dbs. Moreover, the breed-specific UNIQUE SNP-db were further explored and illustrated into three S Tables for the Hereford ([S23 Table](#)), Polish Red ([S24 Table](#)), and Polish HF breeds ([S25 Table](#)), respectively. Finally, Eight putative SNPs from Polish HF were selected from one of the unique SNP-db ([S25 Table](#)), for SNPs validation experiment.

In the following sub-section of SNPs analysis, it was assumed that certain regions of *Bos taurus* genome are still unknown (base = N, according to recent mapping UMD3.1). After applying the both SNP filtering criteria (≥ 10 SNP reads with an accuracy $\geq 90\%$ and $= 100\%$), a total of 1039 best candidates of *de novo* SNPs reads were identified. ([S26](#) and [S27](#) Tables). Lastly, it was also assumed that certain regions of the reference *Bos taurus* genome still contained mutations or errors which were not correctly annotated. Such errors were corrected by analysing the SNPs database through successive annotations. After applying the filtering criteria of ≥ 10 SNP reads with an accuracy $\geq 90\%$, a total of 5.041 potential putative SNPs were identified, whereas, filtering criteria of ≥ 10 SNP reads with an accuracy $= 100\%$, a total of 567 potential putative SNPs was identified ([S28 Table](#)).

Breed-specific SNPs discoveries and QTLs/CGs analysis. For each breed, a set of 76 QTLs/CGs loci (<http://www.animalgenome.org/cgi-bin/QTLdb/index>) was comprehensively investigated ([S29 Table](#)). The chromosomal locations and SNP locations of identified putative SNPs loci for each breed representing three developmental ages were summarized in detail in supplement [S30–S32](#) Tables respectively. This comprehensive QTLs/CGs analysis, allowed us to identify within breed and between breeds putative SNPs for each breed and developmental ages of young bulls ([S33–S35](#) Tables). The results showed that most of the identified top 20 CGs were represented in all three breeds and developmental ages, however, numbers of SNPs hits loci varied across developmental ages. Furthermore, *KCNIP4*, *CCSER1*, *DPP6*, *MAP3K5* and *GHR* genes loci were identified as top 5 potential CGs with highest SNPs hit loci in all three breeds and developmental ages, while *CAST* gene locus was identified as potential CGs expressed with more than 100 SNPs hit in Polish HF and Hereford breeds ([S30–S35](#) Tables). Additionally, a phylogenetic tree [27] was constructed to see how similar the cattle within the same breed. Results revealed that all three breeds were clustered together for each breed and were separated from for each other ([Fig 2](#)). Thus, the study proved to indicate that SNP calling form RNA-seq data has been effective at identified genetic variation.

Breed-specific SNP Validation

In order to validate the breed-specific SNP discovery process, a subset of eight putative SNPs, identified as uniquely specific to the Polish HF breed, were evaluated using single-plex KASP™ assays (LGC Genomics), which are based on fluorescently labelled allele-specific PCR primers. The SNP identity with normal and mutant alleles, genome locations, SNP gene descriptions of selected SNPs are presented in [S36 Table](#). An additional analysis for the estimation of direct relationship between breed and detected SNP polymorphisms was performed using the PROC MIXED SAS 9.2 package with the age as a random effect. Results revealed no significant association between developmental ages and SNPs selected for the validation. However, significant and highly association between breed and SNPs selected for validation were observed ([S37 Table](#)).

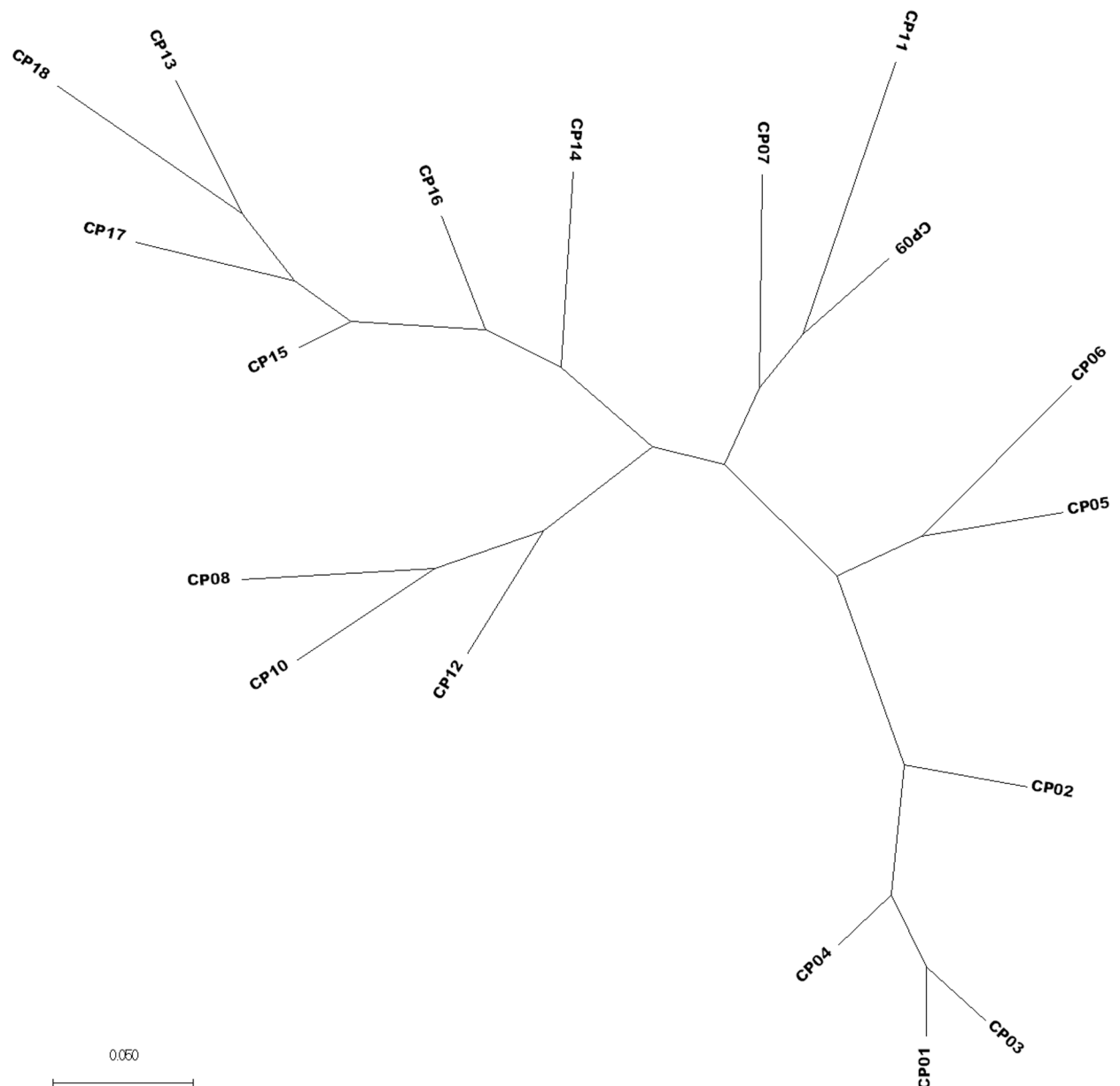


Fig 2. The phylogenetic relationship among RNA-seq samples using Maximum likelihood method. An unrooted phylogeny tree of 18 bulls' samples representing CP01-CP06 (Hereford), CP07-CP12 (Polish Red) and CP13-CP18 (Polish-HF). All nodes were robust at 100% bootstrap support. The scale bar denotes substitutions per site.

doi:10.1371/journal.pone.0161370.g002

Based on this initial analysis, KASP™ SNP assay results were organized according to distributions of genotypes for each validated SNPs locus in all the investigated young bulls from three breeds ([S38 Table](#)). Furthermore, the distribution of and frequencies of SNPs alleles ([S39 Table](#)) results revealed the highest numbers and frequencies of mutant alleles in the SNP loci of all young bulls from the three breeds ($n = 44$), except, BTA4_77555734 and BTA19_27086284 loci. Moreover, the genotypes distributions of each validated SNP were further illustrated specifically for each breed, and are presented in [S40 Table](#). Based on the observed genotypes and allele frequencies ([S38–S40 Tables](#)), the results revealed that the selected SNPs mutations were

polymorphic to tested cattle breed population, and were in accordance with RNA-seq observations. Furthermore, a subset of eight breed-specific putative SNP makers from the Polish HF breed worked well to KASP™ SNP genotyping assay (LGC Genomics), and did not reveal either non-amplification or ambiguous clustering, except in few samples which did not respond to KASP™ assay because of poor DNA quality (Figs 3–10).

Based on the initial KASP™ SNP assays results (S38–S40 Tables), the statistical analysis of SNP validation was carried out by the Genepop software [28, 29]. Using the Fisher's Exact Probability test, the genetic differentiation of SNP alleles and SNP genotypes results revealed significant differences in the SNP alleles frequencies (S41 Table), and SNP genotypes frequencies (S42 Table) among the investigated cattle breeds and SNP loci. Moreover, the comparison of allele and genotypic frequencies of selected SNPs among the investigated cattle breeds showed that highly significant differences between the Polish HF breed vs Polish Red breed, and the Polish HF breed vs Hereford breed, thus confirming the validation of selected Polish HF breed-specific SNP markers. Moreover, using a Markov chain method the SNP validations

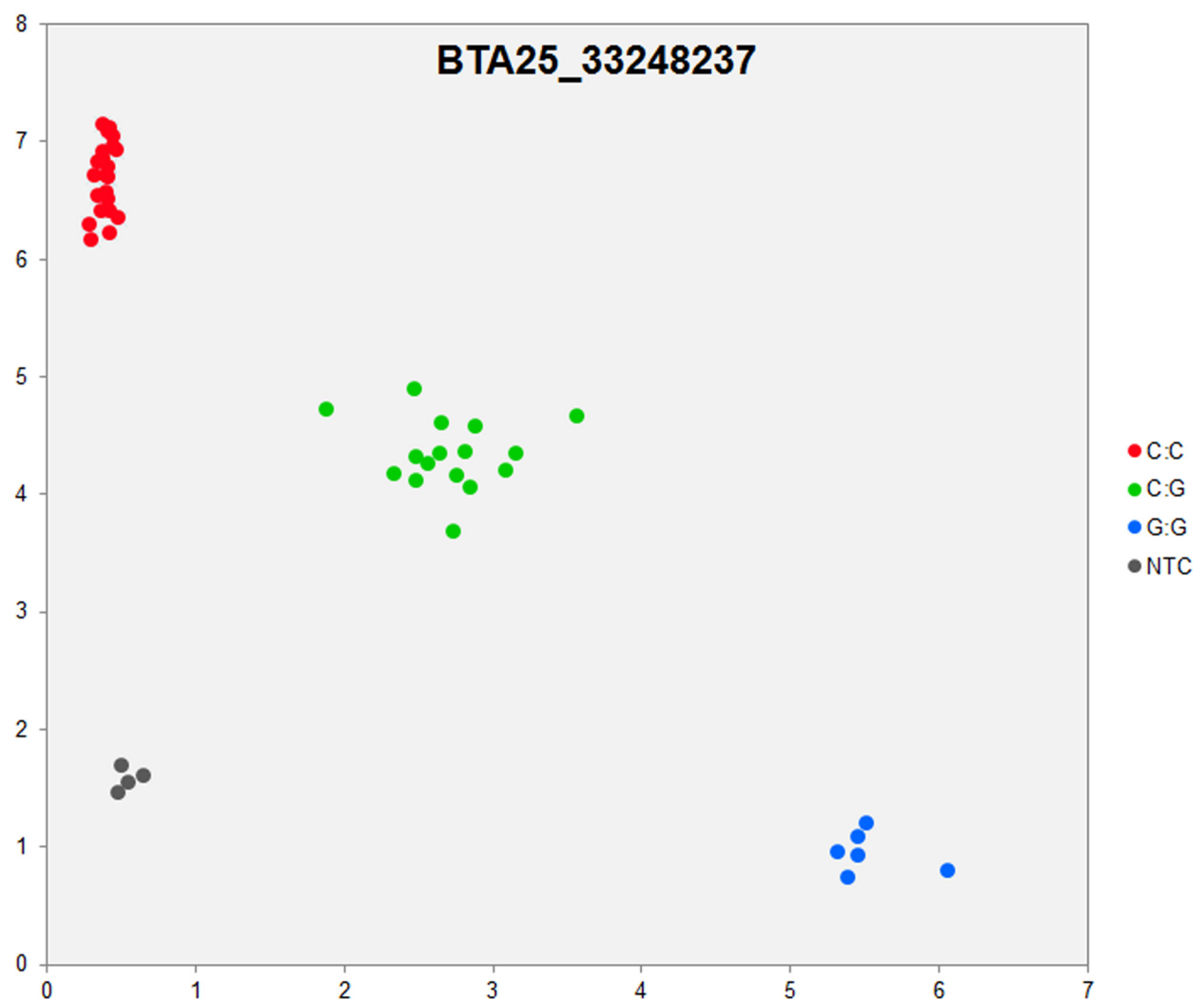


Fig 3. KASP SNP genotyping assay of BTA25_33248237 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g003

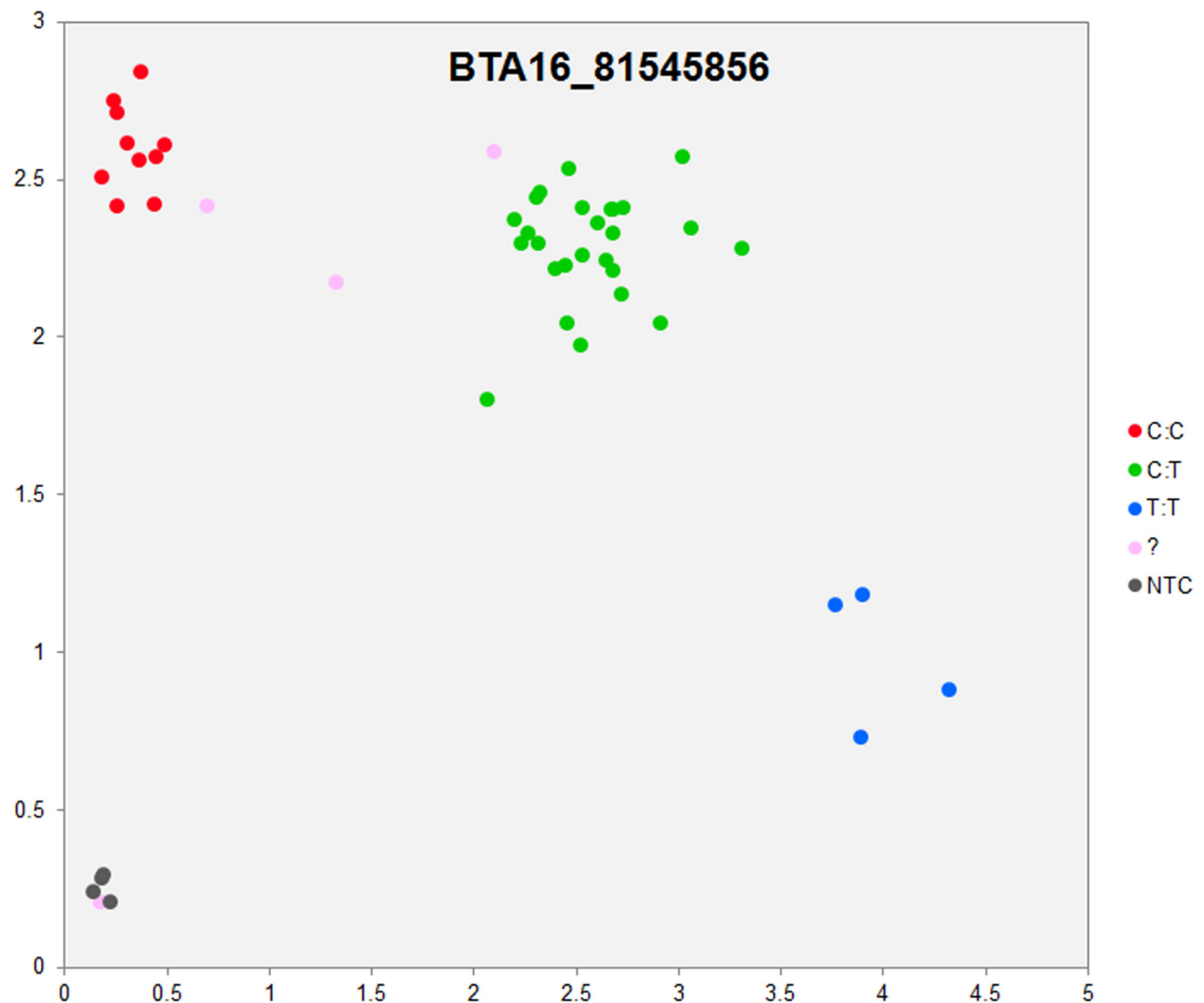


Fig 4. KASP SNP genotyping assay of BTA16_81545856 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g004

results were further confirmed by testing the deviation from Hardy Weinberg equilibrium (HWE) for each SNP locus ([S43 Table](#)), and each breed population ([S44 Table](#)). The results revealed highly significant differences in SNP genotypes and alleles frequencies among the cattle breeds across all eight SNP loci and all three breeds.

Discussion

In cattle genomics, the key challenge in SNP discovery is to distinguish true individual variants (occurring at a low to very low frequencies) from sequencing errors (which often occurs at higher frequencies). Breed-specific SNP discovery is critical for successful implementation of GAS and GS programs for genetics improvement in economically important traits [10–12]. With increasing attention being paid to genomic selection based on the next generation sequencing (NGS) by animal breeders, the primary goal of genome-wide SNP discovery using NGS is to identify large number of markers that provide complete set of genetic variants for

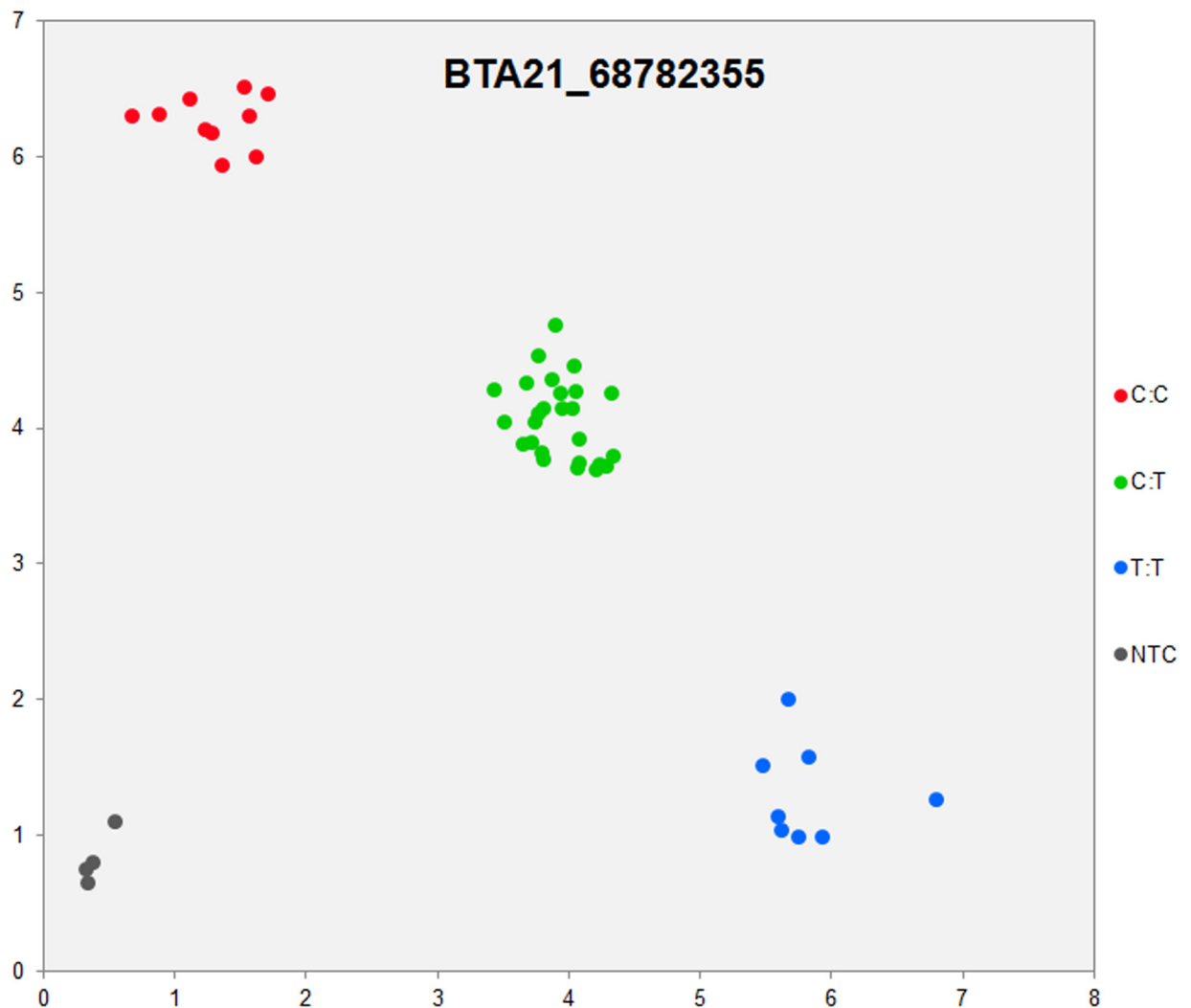


Fig 5. KASP SNP genotyping assay of BTA21_68782355 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g005

economic traits. Herein this paper by using the NGS based RNA-seq strategy, the pituitary gland transcriptome analysis of the three cattle breeds was accomplished by performing the alignment of mapping read, followed by detections of breed-specific SNPs, and finally the validation of SNPs.

Alignment of mapping reads

The mapping of sequence reads was performed with BWA (STAR alignment tool) using their default settings on the same Linux based servers. The study of Wang and colleagues suggested both Bowtie and BWA tools could be utilized for fast and efficient alignment of Illumina short reads [30]. However, other study [31] found that BWA was superior to Bowtie at mapping short reads. For example, a higher percentage of read mappings were achieved using BWA (84.8%) compared to Bowtie (62.3% including the suppressed reads) when Illumina reads of eight genotypes were mapped onto the reference sequence. Therefore, in this paper, the BWA

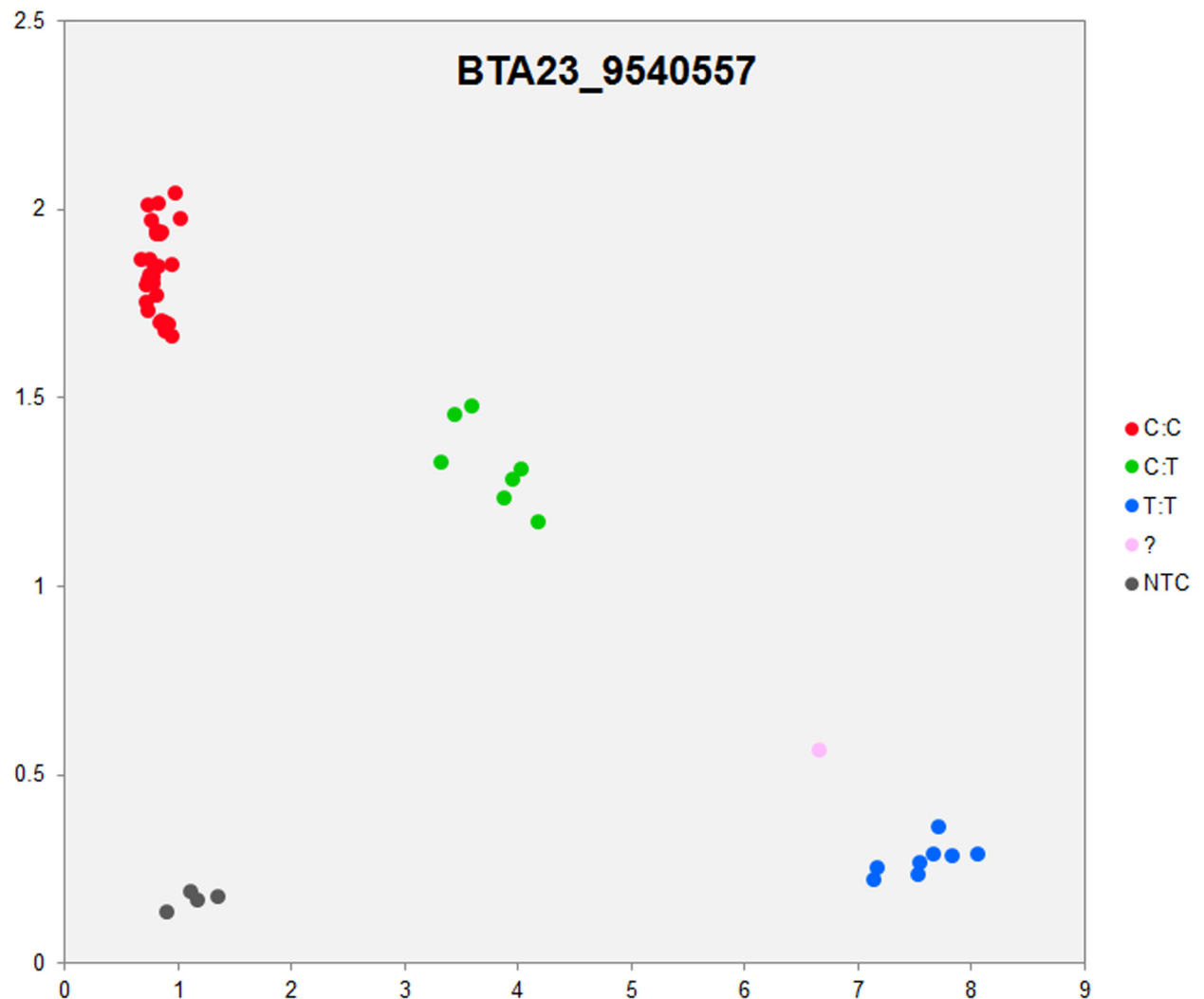


Fig 6. KASP SNP genotyping assay of BTA23_9540557 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g006

based assembly was chosen for downstream analysis because it produced higher percentages of mapped reads.

In cattle genomics, very few tissue-specific RNA-seq experiments on SNP discoveries were reported. According to the PubMed literatures, no RNA-seq experiment on SNP discoveries in bovine pituitary gland transcriptome has been conducted so far, besides our preliminary results [26]. A previous study on the longissimus thoraci transcriptome from three Limousine bull calves revealed approximately 36–46 million paired-end reads per library, and 63% to 67% of the mapped reads were aligned properly to paired end with almost negligible transcriptome contamination, using the BWA program [22].

By sequencing the milk transcriptome mapped uniquely onto the bovine genome, approximately 65% of the RNA-Seq reads were successfully mapped and aligned to the reference genome, and approximately 17,000–19,000 genes were expressed in milk [32]. In another bovine RNA-seq experiment on rumen epithelium tissue, the study of Baldwin and co-workers

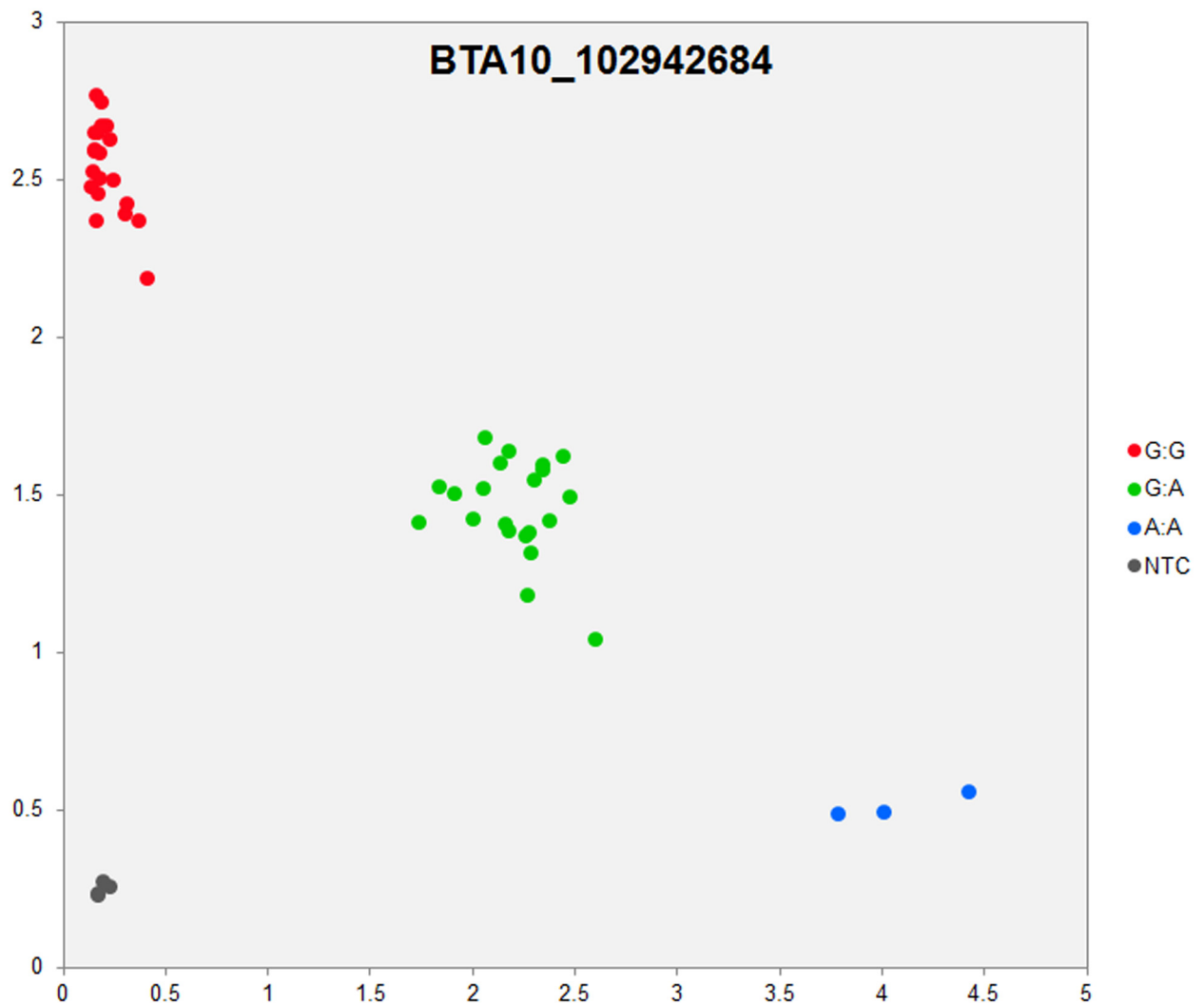


Fig 7. KASP SNP genotyping assay of BTA10_102942684 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g007

found that ~71% of the reads were mapped onto ~17,000 different genes of *Bos taurus* reference genome [33].

A previous study of RNA-seq study on bovine milk transcriptome identified 118 million reads of 36 to 40 bp size, averaging 17 million short-sequence reads for each milk sample. Study further revealed that approximately 70% (82.7 million) reads were properly mapped and remaining 35 million were unmapped. Approximately 87.5% of uniquely mapped reads corresponded to total exon reads, and a small remaining fraction corresponded to total intron reads [16].

While analysing the bovine embryo, the RNA-seq experiment on single bovine blastocysts reported approximately 38 million sequencing reads per embryo, and 9,489 known bovine genes expressed with high correlations of expression levels between the samples. [20]. In another study using multiple tissues (*i.e.*, hypothalamus, pituitary gland, ovary, uterus, and endometrium, longissimus dorsi muscle, adipose, and liver) identified an average of 30 million

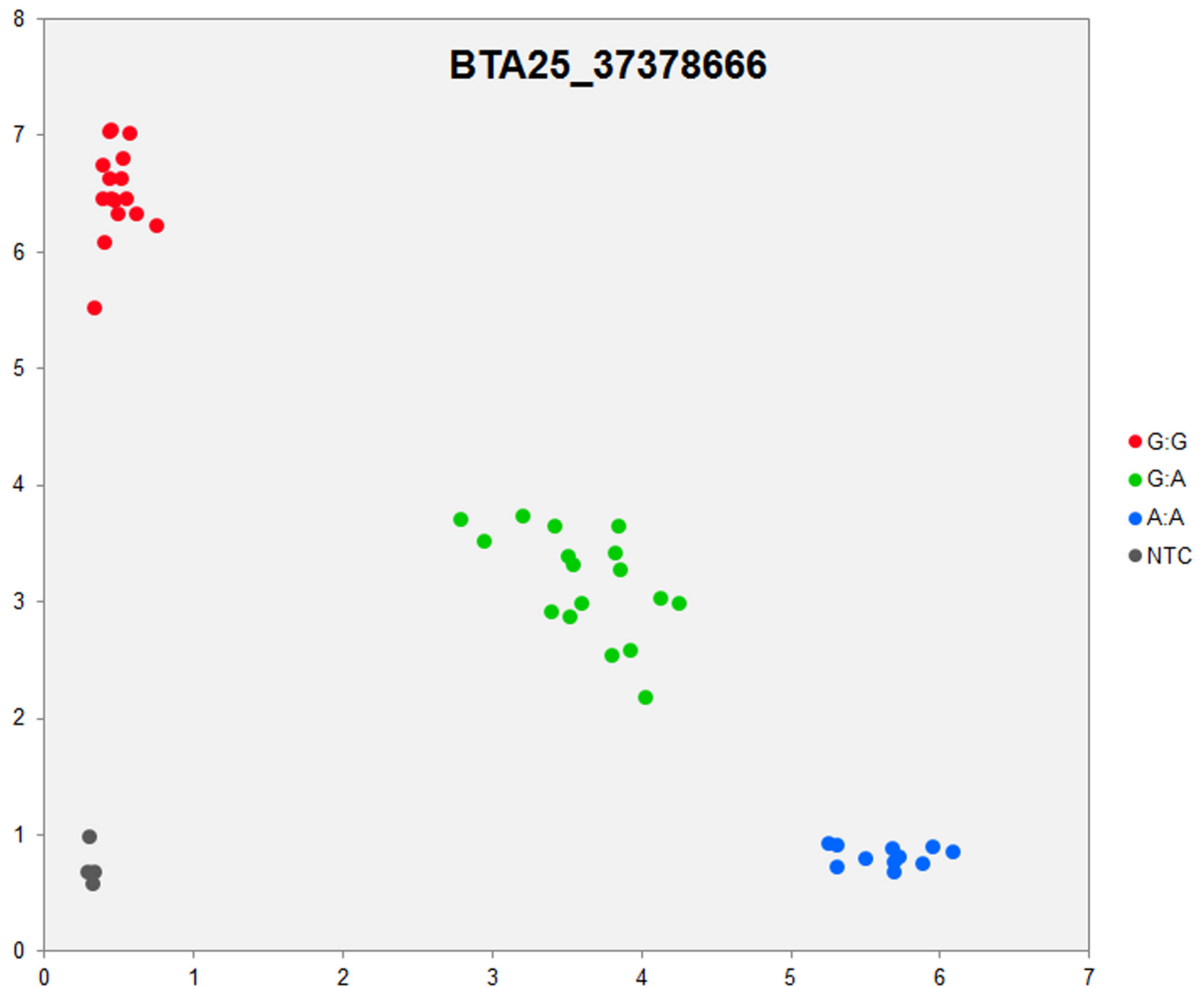


Fig 8. KASP SNP genotyping assay of BTA25_37378666 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g008

sequences reads from each sample tissue, and mapping of approximately 70 to 80% of the sequence reads to the 27,368 annotated genes in the bovine genome assembly UMD3.1.74 [23].

While analysing the RNA-seq experiment of bovine leukocyte transcriptomes from two taurine breeds (Holstein and Jersey), and one indicine breed (Cholistani), a total of approximately 21 million paired-end fragments were sequenced for each of the three breeds. The Tophat results revealed that an approximately 70% of the sequenced fragments were successfully aligned. The study further concluded that more than 90% aligned fragments were mapped to unique genomic regions [18].

Recently, the Qinchuan beef cattle transcriptome analysis in context to longissimus muscle growth and development identified approximately 2 billion pair-end reads of 100 bp in length for bovine longissimus muscle from three embryos and three adult bovines. For Emb135 and 30M samples, approximately 79% and 77% pair-end reads were mapped to the UMD 3.1 reference genome [19].

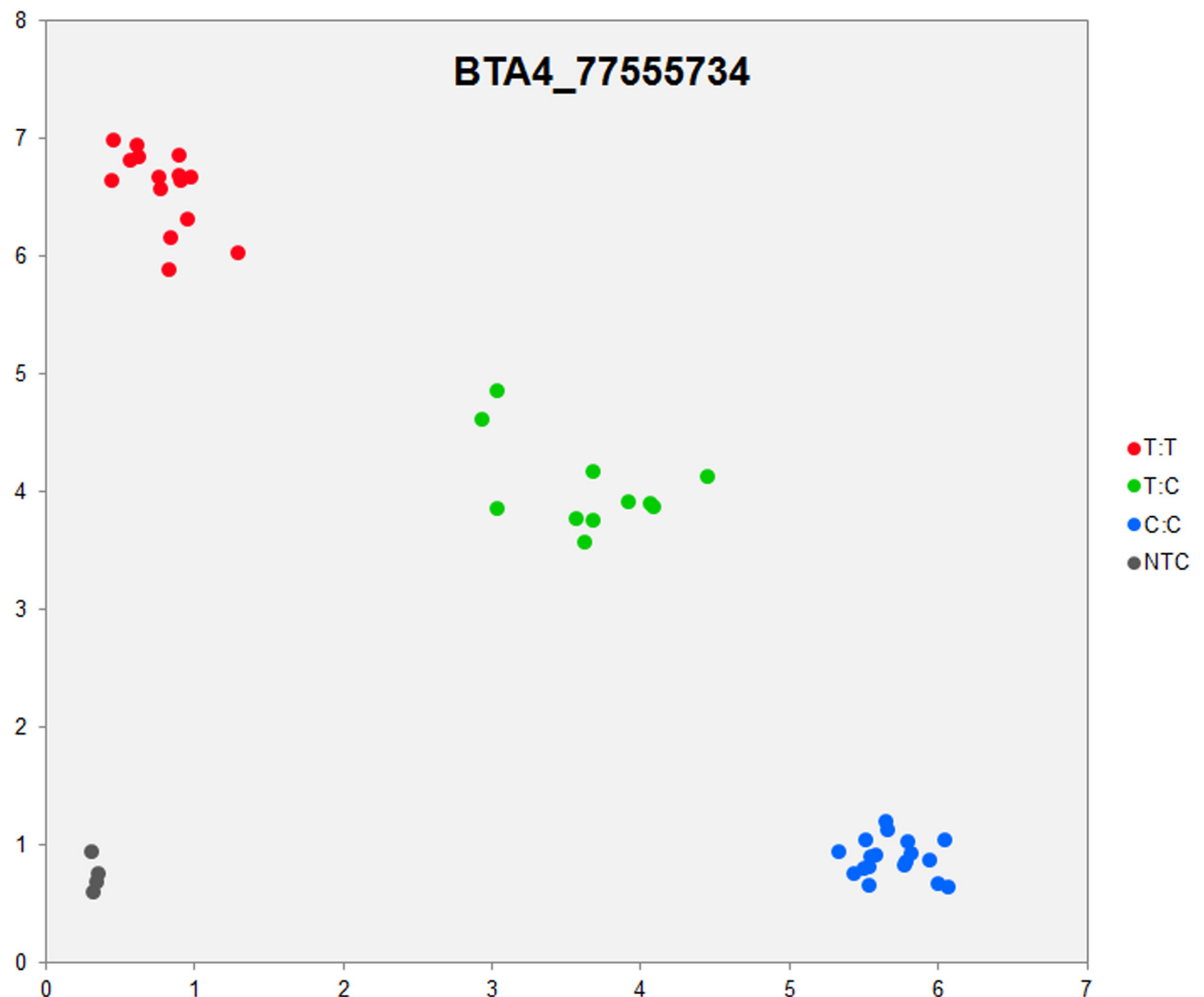


Fig 9. KASP SNP genotyping assay of BTA4_77555734 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g009

Breed-specific SNP discoveries

The alignment file (.gz and.txt format) generated by BWA was used as input files for bovine pituitary gland SNP discovery using SAMtools [34]. The SNP detections were performed using two different stringent parameters of SNP filtering process approaches. The overall constructed SNP database for three breeds yielded approx. 13.7 million of SNPs. While, after filtering to different stringent parameters, 0.59 million and 0.49 million SNPs were detected for the stringent parameters of with minimum 10 SNP reads per base and with an accuracy of $\geq 90\%$ and $= 100\%$, respectively.

In an analysis of the transcriptome, utilizing RNA-seq of bovine Longissimus thoraci from three Limousine bull calves a total of 34,376 different SNP were detected [22]. Amongst these SNPs, 8,974 (26%) were homozygous in all three sequenced samples, corresponding presumably to differences between Limousine and Hereford. This study also found that there were

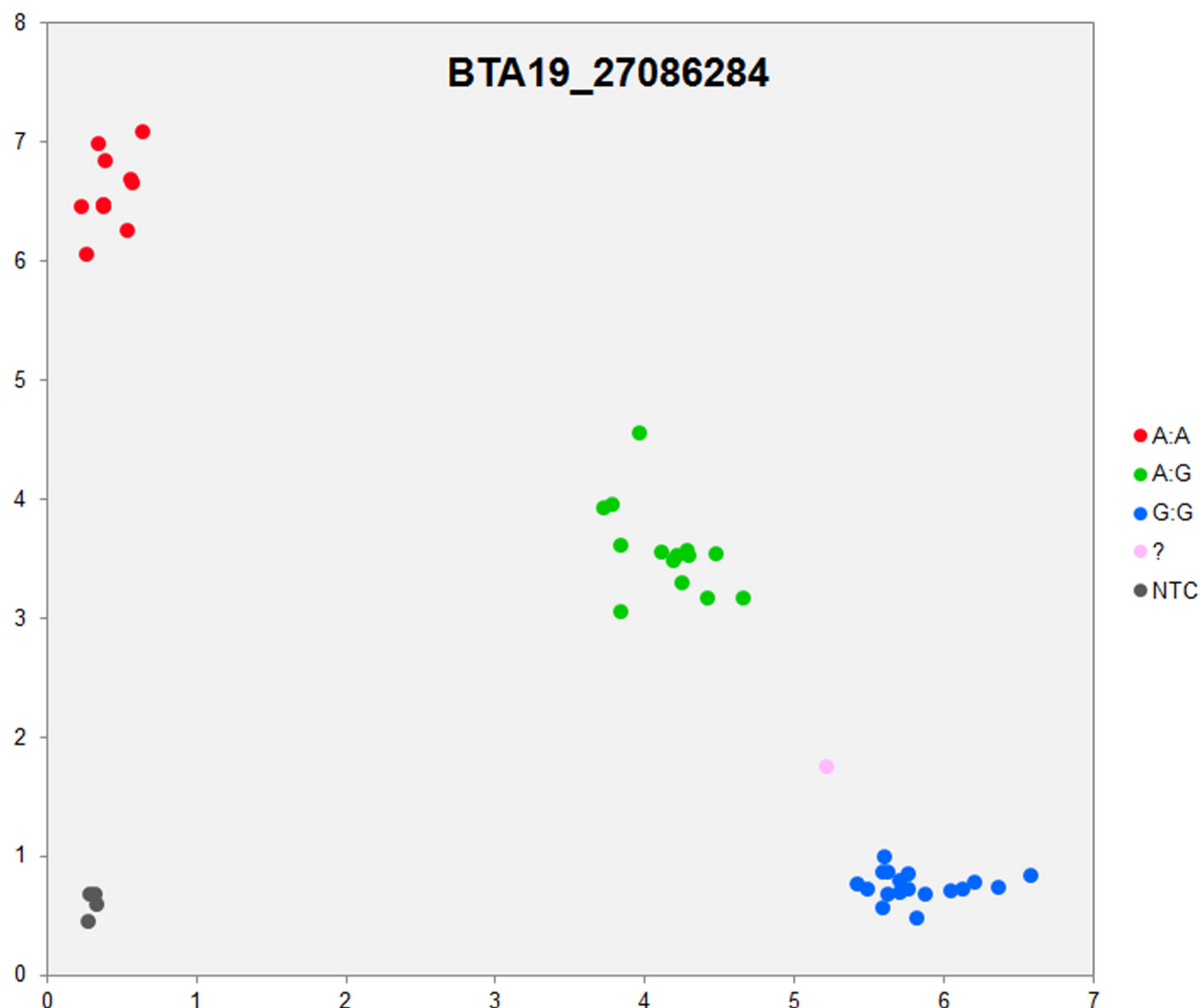


Fig 10. KASP SNP genotyping assay of BTA19_27086284 locus showing the data for single KASP assays on a single cluster plot.

doi:10.1371/journal.pone.0161370.g010

30,998 bi-allelic SNPs mapping to coding regions, 38.6% of which were previously found and recorded in dBs [22].

In a study on milk transcriptome, approximately 100,000 SNPs located in genes expressed in milk samples from the Holstein cows were identified [16]. However, only 32% of SNPs (33,045) were polymorphic within their seven Holstein cows. In another study, [23], putative SNPs within three investigated cattle breeds *i.e.*, the Holstein, Jersey, and Cholistani were reported. Results showed that among 22,100,344 surveyed bases in Holstein, 32,547 (0.147%) were reliably called SNPs, and 39,370 out of 19,967,581 (0.197%) bases were polymorphic within Cholistani. Using different SNPs variant calling tools, a total of 31,334 and 30,618 putative SNPs were identified for the longissimus pooled samples of adults and pooled samples of embryo. For both samples, the SNP detection stringency conditions of i) at least two unique mapping reads that support the polymorphic nucleotide, and ii) a quality score of ≥ 20 were utilized [19].

The potential applications of RNA-Seq present unique benefits in terms of SNP analysis because of its wide dynamic range and ability to identify functional sequence variants. The detection and categorization of SNPs within animal production systems have been performed extensively [15]. The use of these genetic variants as markers and predictors of performance in a large variety of economic traits (*i.e.* meat production, milk production, fertility, calving ease, etc.) is quite common in animal breeding practices [35–37]. In the context of early body growth and development trait, classifying SNPs between samples of varying viability, ages, or breed allows for discovery of novel markers of growth traits and characterization of critical regulatory mechanisms of body growth and development. Traditionally, the use of transcript sequence for SNP variant detection has been performed with various assays [38], and more recently in cattle using RNA sequencing data [16].

Conclusions

This is the first study to report breed-specific SNPs discovery using NGS based RNA-seq in bovine pituitary gland. The study generated nearly fourteen million breed-specific SNP-db based on the expressed genes of the bovine pituitary gland, with mapping accuracy comparable or better to previous works in *Bos taurus*. These databases of breed-specific SNP as well as identified putative SNPs within the QTLs/CGs for bovine body growth and development trait will improve the genomic resources available for cattle, especially for beef breeds, and may also prove useful to study the mechanisms underlying the genetic variability in meat quality and reproduction traits. KASP™ SNP assay was proven to be an efficient cost effective method to validate the breed specific putative SNPs originating from RNA-seq experiments.

Materials and Methods

Animals

The bovine pituitary gland samples from young growing bulls were collected from Institute of Genetics and Animal Breeding, Polish Academy of Science (PAS), Jastrzębiec, Poland. After slaughtering of the animals, the collected tissues were immediately kept in liquid nitrogen, and finally stored in deep freezer at -80°C . All procedures involving animals were performed in accordance with the guiding principles for the care and use of research animals, and were approved by the local ethics commission (permission No. 3/2005) of Institute of Genetics and Animal Breeding, Polish Academy of Science (PAS), Jastrzębiec, Poland.

The experimental design comprised of 18 young bulls for RNA-seq experiment, and 44 young growing bulls for SNP validation experiment (Table 5), in a panel of three selected cattle

Table 5. The distribution (n) of investigating young bulls in a breed-specific experimental design representing bovine pituitary gland transcriptome.

Breeds	6 months	9 months	12 months	Total
RNA-seq experiment				
Polish HF	2	2	2	6
Polish Red	2	2	2	6
Total	6	6	6	18
SNP validation experiment				
Hereford	2	2	2	6
Polish HF	2	2	2	6
Polish Red	2	2	2	6
Total	6	6	6	18

doi:10.1371/journal.pone.0161370.t005

breeds: Polish HF, Polish Red, and Hereford, respectively. All experimental animals were reared at Institute of Genetics and Animal Breeding, Jastrzębiec, Poland in a closed herd, and providing uniform feeding and environmental conditions.

Laboratory procedures

Total RNA was extracted and prepared from 50–60 mg of frozen bovine pituitary gland tissues ($n = 18$) using the guanidinium thiocyanate method [39] (TRIzol reagent: Invitrogen, Carlsbad, CA, USA). Preliminary RNA samples were evaluated with the Agilent BioAnalyzer using the Nano RNA Kit. Only samples with RNA Integrity Number (RIN) > 7.0 , and 5 total μg were used for library preparation, and two biological replicates were used for each age/breed group. The mRNA isolation was carried out by using the Dynabeads[®] mRNA Direct[™] kit (Thermo Fisher), and following recovery, polyA RNA was evaluated using the BioAnalyzer Nano mRNA assay, verify that samples contain less than 2.0% rRNA. For dUTP directional mRNA libraries preparation, 25–100ng mRNA was fragmented by chemical hydrolysis, converted to first strand cDNA with random hexamers, and second strand synthesized with dUTP according to the NEBNext Ultra Directional RNA library preparation Kit for Illumina (New England Bio Labs). The cDNA fragments were end-repaired, A-tailed, and ligated to the TruSeq y-tail single indexes from Illumina TruSeq DNA kit. Indexed libraries were cut with USER enzyme, and PCR amplified for 12 cycles. Finally, to achieve the highest quality data on Illumina sequencing platforms, optimum cluster deposition was made by quantitation of libraries using qPCR according to the Illumina Sequencing Library qPCR Quantification Guide (Kapa Biosciences). 156x156 bp paired-end sequence reads were generated using the Illumina NextSeq 500 platform High Output/300 cycle kits from Illumina.

Bioinformatics analysis

Pre-processing analysis of RNA-seq data and sequence quality control (QC) was performed prior to read alignment to reference genome. For all libraries, the adaptor sequences were removed using the cutadapt software with minimum overlap length was set to 10 and error rate was set to 0.05 [40]. After adaptor trimming, the low quality bases were trimmed from 3'-end followed by sequence quality control (FastQC). The post processing analysis of RNA-seq data was performed i) to map the sequencing read (read alignment), as well as ii) to discover the breed-specific putative SNPs.

1. **Alignment of RNA-seq sequencing reads to the reference genome.** The pre-processed short paired-end reads were aligned to the reference genome Ensembl75_UMD3-1.1 plus the Chromosome Y from Btau_4.6.1 assembly, by using BWA version 0.7.5-r404 [41]. The HT-Seq framework, version 0.5.3p9 (<https://pypi.python.org/pypi/HTSeq/0.5.3p7>), was used to count the aligned reads in genes using the STAR BWA tools [42].
2. **Breed-specific discoveries of novel putative SNPs in Polish HF, Polish Red, and Hereford cattle.** For the analysis of SNPs discoveries (variant calling), the SAMtools mPileUp package to call SNPs and *indels* [34] was utilized to detect the putative breed-specific SNPs in bovine pituitary gland transcriptome. Using Microsoft Office Excel, the following two stringent parameters of SNP filtering were performed.
 - Stringent parameters of SNP filtering with minimum 10 SNP reads per base and with an accuracy of $\geq 90\%$.

- Stringent parameters of SNP filtering with minimum 10 SNP reads per base and with an accuracy = 100%.
Similarly, using Microsoft Office Excel the breed-specific SNP database of bovine pituitary gland were further trimmed to a highly reliable breed-specific SNP-db according to SNP type, *de novo* SNP and annotated SNPs, respectively.
- 3. **Breed-specific SNPs discoveries and QTLs/CGs analysis.** The publicly available animal quantitative trait loci (QTL) database (Animal QTLdb: <http://www.animalgenome.org/cgi-bin/QTLdb/index>) representing trait associated QTLs as well as candidate gene and association data (GWAS) mapped to bovine genomes was utilized to find putative SNPs within the candidate genes for bovine body growth and development trait [43]. A total of 76 potential QTLs/CGs (S29 Table) for bovine body growth and developmental trait was selected to investigate the RNA-seq SNP-db. To identify the putative SNPs within the CGs loci, stringent SNP filtering was carried out on both RNA-seq and bovine QTL databases using Microsoft Office Excel.
- 4. **Phylogenetic analysis.** The phylogenetic package [27] to call SNPs and *indels* was utilized to detect the putative breed-specific SNPs in bovine pituitary gland transcriptome. Phylogenetic analysis was based on the combined transcriptome datasets of all three breeds using Maximum likelihood method. An unrooted phylogeny tree of 18 bull's samples representing CP01-CP06 (Hereford), CP07-CP12 (Polish Red) and CP13-CP18 (Polish-HF) was constructed by snp-phylo using SAMtools unified SNPs with following parameter “-c 3 -m 0.4 -a 30 -P CP1-18.snphylo -l 0.2 -m 0.2 -M 0.2”.

KASP™ assays design and SNP validation. Eight breed-specific SNPs (Polish HF) were selected for KASP™ SNP assays validation experiment. The primer sequences are presented in Table 6. The genomic DNA of 44 young bulls from Polish HF, Polish Red, and Hereford were isolated from pituitary gland tissue using the MasterPure™ DNA purification kit with some modifications (Epicentre). All 44 DNA samples were then shipped to LGC genomics Teddington, Middlesex, TW11-0LY, UK (<http://www.lgcgroup.com>) to perform the KASP™ SNP genotyping assay including primer design. For each of selected breed-specific SNP marker, two allele specific forward primers and a common reverse primer were designed for use in fluorescence based competitive allele-specific PCR assays. Primers designed were done by LGC genomics, UK, by blasting the region of +200bp and -200pb around the nucleotide variation using the PrimerPicker [44]. The KASP™ genotyping reaction and PCR thermocycling conditions were carried out according to manufacturer's recommendations (LGC Genomics, UK). PCR reaction mixture was carried out in a 10 µl-volume consisting of 2.5 µl DNA (10 ng/µl), 2.5 µl

Table 6. SNP-ID and primer sequences of selected breed-specific SNPs of Polish HF cattle.

SNP-ID	SNP assay primer sequences
BTA25_33248237	TCAACTTATGGAAAT[G/C]AGTCAACAGAAGAG
BTA16_81545856	GAGGGCTGGGGGACC[T/C]GGTCGGGAGGCAGC
BTA21_68782355	GTGTCAAGATTACCT[T/C]CGTAGTTCCTGGACC
BTA23_9540557	GCACATCTCAAGCAT[T/C]GAGATTTATGATCC
BTA10_102942684	GATCAGAGACTGCCQ[A/G]AGATCGTCAAACAA
BTA25_37378666	CTCAAATCTTGAAAG[A/G]AGTGTAAAGACATAA
BTA4_77555734	TCAGCCTTCTACTCQ[C/T]TCAGGCAGCCAGGG
BTA19_27086284	TCCTTTTATGTACGG[G/A]TTAAAAAGGAAAG

doi:10.1371/journal.pone.0161370.t006

2X KASP™ Master mix v4.0, and 0.07 µl 72X KASP™ Primer mix. PCR amplification and fluorescent end-point genotyping was carried out in a LightCycler thermocycler. Finally, data was analysed using Kluster-caller software (LGC Genomics. UK) to identify SNP genotypes.

Statistical analysis. Using the Genepop software (<http://genepop.curtin.edu.au/>), the following statistical tests were carried out to validate the breed-specific SNPs.

1. Estimates of direct relationship between breed and detected SNP polymorphisms was performed by the use PROC MIXED SAS 9.2 package, with the age as a random effect.
2. A Fisher test was performed across all populations, and for each locus pair to compute an unbiased estimate of the P-value.
3. Genic and genotypic differentiation was tested for each locus for all populations, and for each population pair across all loci.
4. Significance of differences in allele frequencies between samples were determined by the Exact Fisher's method and the Markov Chain procedure [28], and the Fisher combined test was computed as a global test over loci to determine the overall significance.
5. Deviation from Hardy-Weinberg equilibrium for each population was calculated by a probability exact test using a Markov Chain method [29] as implemented in Genepop.

Supporting Information

S1 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-1 of Hereford cattle aged 6 months.

(XLSX)

S2 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-2 of Hereford cattle aged 6 months.

(XLSX)

S3 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-3 of Hereford cattle aged 9 months.

(XLSX)

S4 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-4 of Hereford cattle aged 9 months.

(XLSX)

S5 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-5 of Hereford cattle aged 12 months.

(XLSX)

S6 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-6 of Hereford cattle aged 12 months.

(XLSX)

S7 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-7 of Polish Red cattle aged 6 months.

(XLSX)

S8 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-8 of Polish Red cattle aged 6 months.

(XLSX)

S9 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-9 of Polish Red cattle aged 9 months.

(XLSX)

S10 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-10 of Polish Red cattle aged 9 months.

(XLSX)

S11 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-11 of Polish Red cattle aged 12 months.

(XLSX)

S12 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-12 of Polish Red cattle aged 12 months.

(XLSX)

S13 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-13 of Polish HF cattle aged 6 months.

(XLSX)

S14 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-14 of Polish HF cattle aged 6 months.

(XLSX)

S15 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-15 of Polish HF cattle aged 9 months.

(XLSX)

S16 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-16 of Polish HF cattle aged 9 months.

(XLSX)

S17 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-17 of Polish HF cattle aged 12 months.

(XLSX)

S18 Table. RNA-seq SNP-db of bovine pituitary gland tissue of young bull-18 of Polish HF cattle aged 12 months.

(XLSX)

S19 Table. Construction of filtered/trimmed breed-specific SNP-db of bovine pituitary gland transcriptome.

(XLSX)

S20 Table. Analysis of breed-specific SNP-db using the SNP filtering criteria of a minimum 10 SNP reads per base with an accuracy $\geq 90\%$.

(XLSX)

S21 Table. Analysis of breed-specific SNP-db using the SNP filtering criteria of a minimum 10 SNP reads per base with an accuracy = 100%.

(XLSX)

S22 Table. Screening of SNP type and occurrence in the experimental breed-specific SNP-db.

(XLSX)

S23 Table. Identification of unique SNPs using 90% and 100% SNP filtering in Hereford cattle.

(XLSX)

S24 Table. Identification of unique SNPs using 90% and 100% SNP filtering in Polish Red cattle.

(XLSX)

S25 Table. Identification of unique SNPs using 90% and 100% SNP filtering in Polish HF cattle.

(XLSX)

S26 Table. Identification de novo SNPs in Hereford, Polish Red and Polish HF cattle using 90% and 100% SNP filtering criteria.

(XLSX)

S27 Table. Distribution of Identified de novo SNPs counts in the investigated young bulls on the basis of [S26 Table](#).

(XLSX)

S28 Table. Annotation of SNPs in Hereford, Polish Red and Polish HF cattle.

(XLSX)

S29 Table. Lists of 76 CGs representing full names, UMD3.1 genome locations, chromosomal locations and web links at bovine QTL-DB.

(XLSX)

S30 Table. Identification of putative SNPs hits of RNA-seq data on 76 potential candidate genes from bovine QTL-db in Polish Red cattle.

(XLSM)

S31 Table. Identification of putative SNPs hits of RNA-seq data on 76 potential candidate genes from bovine QTL-db in Polish HF cattle.

(XLSX)

S32 Table. Identification of putative SNPs hits of RNA-seq data on 76 potential candidate genes from bovine QTL-db in Hereford cattle.

(XLSX)

S33 Table. Summary of identified putative SNPs hits of RNA-seq data on 76 potential candidate genes from bovine QTL-db in Polish Red cattle.

(XLSM)

S34 Table. Summary of identified putative SNPs hits of RNA-seq data on 76 potential candidate genes from bovine QTL-db in Polish HF cattle.

(XLSM)

S35 Table. Summary of identified putative SNPs hits of RNA-seq data on 76 potential candidate genes from bovine QTL-db in Hereford cattle.

(XLSX)

S36 Table. SNP-ID, genome locations, SNP mutations, SNP genes descriptions of selected SNPs specific to Polish HF breed in RNA-seq experiment.

(XLSX)

S37 Table. Fixed effect of breeds and developmental ages on validated SNPs markers Using REML mixed model procedure.

(XLSX)

S38 Table. Distribution of SNP genotypes in the investigated young bulls.

(XLSX)

S39 Table. Distribution and frequencies of SNP alleles in the investigated young bulls.

(XLSX)

S40 Table. Breed specific distribution of SNP genotypes in the investigated young bulls.

(XLSX)

S41 Table. Genetic differentiation of SNP alleles frequencies among investigated cattle breeds using the Fisher's exact Probability test.

(XLSX)

S42 Table. Genetic differentiation of SNP genotypes frequencies among investigated cattle breeds using the Fisher's exact Probability test.

(XLSX)

S43 Table. Hardy-Weinberg test for genetic differentiation of investigated SNP loci using the Markov chain method.

(XLSX)

S44 Table. Hardy-Weinberg test for genetic differentiation of investigated cattle breeds using the Markov chain method.

(XLSX)

Acknowledgments

Authors greatly thankful to Institute of Genetics and Animal Breeding of the Polish Academy of Sciences, Jastrzebiec, Poland, for providing the pituitary gland tissue materials essential for the study. Authors also thank to LGC genomics, UK for performing the SNP validation experiment, and Clinscript llc, NJ, USA for professional proofreading and editing services to manuscript.

Author Contributions

Conceived and designed the experiments: CSP HNK TG JH LZ UC LF PS KW.

Performed the experiments: CSP RS PD PB MS PW BG.

Analyzed the data: MP AS MO YF DK.

Contributed reagents/materials/analysis tools: CSP RS PD PB PW BG.

Wrote the paper: CSP HNK TG JH LZ UC LF PS KW.

References

1. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 1–8. PMID: [18175409](#)
2. Pérez-Enciso M, Ferretti L. (2010). Massive parallel sequencing in animal genetics: wherefroms and wheretos. *Anim Genet* 41: 561–9. doi: [10.1111/j.1365-2052.2010.02057.x](#) PMID: [20477787](#)
3. Qian X, Ba Y, Zhuang Q, Zhong G. (2014). RNA-Seq technology and its application in fish transcriptomics. *OMICS* 18: 98–110. doi: [10.1089/omi.2013.0110](#) PMID: [24380445](#)

4. Pareek CS, Smoczynski R, Tretyn A. (2011). Sequencing technologies and genome sequencing. *J Appl Genet* 52: 413–35. doi: [10.1007/s13353-011-0057-x](https://doi.org/10.1007/s13353-011-0057-x) PMID: [21698376](https://pubmed.ncbi.nlm.nih.gov/21698376/)
5. Elsik CG, Tellam RL, Worley KC. (2009). Bovine Genome Sequencing and Analysis Consortium. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324: 522–528. doi: [10.1126/science.1169588](https://doi.org/10.1126/science.1169588) PMID: [19390049](https://pubmed.ncbi.nlm.nih.gov/19390049/)
6. Bovine HapMap Consortium, Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324: 528–32. doi: [10.1126/science.1167936](https://doi.org/10.1126/science.1167936) PMID: [19390050](https://pubmed.ncbi.nlm.nih.gov/19390050/)
7. Pareek CS, Smoczynski R, Pierzchala M, Czarnik U, Tretyn A. (2011). From genotype to phenotype in bovine functional genomics. *Brief Funct Genomics* 10: 165–71. doi: [10.1093/bfpg/elr019](https://doi.org/10.1093/bfpg/elr019) PMID: [21628315](https://pubmed.ncbi.nlm.nih.gov/21628315/)
8. Zhu M, Zhao S. (2007). Candidate gene identification approach: progress and challenges. *Int J Biol Sci* 3: 420–427. PMID: [17998950](https://pubmed.ncbi.nlm.nih.gov/17998950/)
9. Li C, Basarab J, Snelling W, Benkel B, Murdoch B, et al. (2004). Assessment of positional candidate genes MYF5 and IGF1 for growth on bovine chromosome 5 in commercial lines of *Bos taurus*. *J Anim Sci* 82: 1–7.
10. Lindholm-Perry AK, Kuehn LA, Smith TP, Ferrell CL, Jenkins TG, et al. (2012). A region on BTA14 that includes the Positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. *Anim Genet* 43: 216–9. doi: [10.1111/j.1365-2052.2011.02232.x](https://doi.org/10.1111/j.1365-2052.2011.02232.x) PMID: [22404358](https://pubmed.ncbi.nlm.nih.gov/22404358/)
11. Morsci NS, Schnabel RD, Taylor JF. (2006). Association analysis of adiponectin and somatostatin polymorphisms on BTA1 with growth and carcass traits in Angus cattle. *Anim Genet* 37: 554–62. PMID: [17121600](https://pubmed.ncbi.nlm.nih.gov/17121600/)
12. Bai Y, Sartor M, Cavalcoti J. (2012). Current status and future perspectives for sequencing livestock genomes. *J Anim Sci Biotechnol* 3: 8. doi: [10.1186/2049-1891-3-8](https://doi.org/10.1186/2049-1891-3-8) PMID: [22958500](https://pubmed.ncbi.nlm.nih.gov/22958500/)
13. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, et al. (1995). Mapping quantitative trait loci controlling milk production by exploiting progeny testing. *Genetics* 139: 907–20. PMID: [7713441](https://pubmed.ncbi.nlm.nih.gov/7713441/)
14. Dekkers JC. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim Sci* 82: E313–28. PMID: [15471812](https://pubmed.ncbi.nlm.nih.gov/15471812/)
15. Hayes B, Goddard M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome* 53: 876–83. doi: [10.1139/G10-076](https://doi.org/10.1139/G10-076) PMID: [21076503](https://pubmed.ncbi.nlm.nih.gov/21076503/)
16. Canovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano JF. (2010). SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome* 21: 592–598. doi: [10.1007/s00335-010-9297-z](https://doi.org/10.1007/s00335-010-9297-z) PMID: [21057797](https://pubmed.ncbi.nlm.nih.gov/21057797/)
17. Fortes MR, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, et al. (2012). Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *J Anim Sci* 90: 2894–906. doi: [10.2527/jas.2011-4601](https://doi.org/10.2527/jas.2011-4601) PMID: [22739780](https://pubmed.ncbi.nlm.nih.gov/22739780/)
18. Huang W, Nadeem A, Zhang B, Babar M, Solter M, et al. (2012). Characterization and comparison of the leukocyte transcriptomes of three cattle breeds. *PLoS ONE* 7: e30244. doi: [10.1371/journal.pone.0030244](https://doi.org/10.1371/journal.pone.0030244) PMID: [22291923](https://pubmed.ncbi.nlm.nih.gov/22291923/)
19. He H, Liu X. (2013). Characterization of Transcriptional Complexity during Longissimus Muscle Development in Bovines Using High-Throughput Sequencing. *PLoS One* 8: e64356. doi: [10.1371/journal.pone.0064356](https://doi.org/10.1371/journal.pone.0064356) PMID: [23762238](https://pubmed.ncbi.nlm.nih.gov/23762238/)
20. Chitwood JL, Rincon G, Kaiser GG, Medrano JF, Ross PJ. (2013). RNA-seq analysis of single bovine blastocysts. *BMC Genomics* 14: 350. doi: [10.1186/1471-2164-14-350](https://doi.org/10.1186/1471-2164-14-350) PMID: [23705625](https://pubmed.ncbi.nlm.nih.gov/23705625/)
21. Koringa PG, Jakhesara SJ, Bhatt VD, Patel AB, Dash D, et al. (2013). Transcriptome analysis and SNP identification in SCC of horn in (*Bos indicus*) Indian cattle. *Gene* 530: 119–26. doi: [10.1016/j.gene.2013.07.061](https://doi.org/10.1016/j.gene.2013.07.061) PMID: [23978612](https://pubmed.ncbi.nlm.nih.gov/23978612/)
22. Djari A, Esquerré D, Weiss B, Martins F, Meersseman C, et al. (2013). Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics* 14: 307. doi: [10.1186/1471-2164-14-307](https://doi.org/10.1186/1471-2164-14-307) PMID: [23651547](https://pubmed.ncbi.nlm.nih.gov/23651547/)
23. Cánovas A, Reverter A, DeAtley KL, Ashley RL, Colgrave ML, et al. (2014). Multi-Tissue Omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PLoS One* 9: e102551. doi: [10.1371/journal.pone.0102551](https://doi.org/10.1371/journal.pone.0102551) PMID: [25048735](https://pubmed.ncbi.nlm.nih.gov/25048735/)
24. Zhou Y, Sun J, Li C, Wang Y, Li L, et al. (2014). Characterization of transcriptional complexity during adipose tissue development in bovines of different ages and sexes. *PLoS One* 9: e113881.

25. Pauciuolo A, Küpper J, Brandt H, Donat K, Iannuzzi L, et al. (2015). Wingless-type MMTV integration site family member 2 (WNT2) gene is associated with resistance to MAP in faecal culture and antibody response in Holstein cattle. *Anim Genet* 46: 122–32. doi: [10.1111/age.12261](https://doi.org/10.1111/age.12261) PMID: [25643727](https://pubmed.ncbi.nlm.nih.gov/25643727/)
26. Pareek CS, Smoczyński R, Dziuba P, Blaszczyk P, Trzaska M, et al. (2016). Identification of Putative SNPs in Bovine Pituitary Gland Transcriptome using RNA-seq. *Proceedings of PAG XIV*, Jan. 09–14, 2016, San Diego, CA, USA.
27. Lee TH, Guo H, Wang X, Kim C, Paterson AH. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, 15: 162. doi: [10.1186/1471-2164-15-162](https://doi.org/10.1186/1471-2164-15-162) PMID: [24571581](https://pubmed.ncbi.nlm.nih.gov/24571581/)
28. Raymond M, Rousset F. (1995). Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86: 248–49.
29. Guo SW, Thompson EA. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361–72. PMID: [1637966](https://pubmed.ncbi.nlm.nih.gov/1637966/)
30. Wang W, Wei Z, Lam T-W, Wang J. (2011). Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Science Reports* 1:55.
31. Kumar S, You FM, Cloutier S. (2012). Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. *BMC Genomics* 13: 684. doi: [10.1186/1471-2164-13-684](https://doi.org/10.1186/1471-2164-13-684) PMID: [23216845](https://pubmed.ncbi.nlm.nih.gov/23216845/)
32. Wickramasinghe S, Rincon G, Islas-Trejo A, Medrano JF. (2012). Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* 13: 45. doi: [10.1186/1471-2164-13-45](https://doi.org/10.1186/1471-2164-13-45) PMID: [22276848](https://pubmed.ncbi.nlm.nih.gov/22276848/)
33. Baldwin RL 6th, Wu S, Li W, Li C, Bequette BJ, et al. (2012). Quantification of transcriptome responses of the rumen epithelium to butyrate infusion using RNA-seq technology. *Gene Regul Syst Bio* 6: 67–80. doi: [10.4137/GRSB.S9687](https://doi.org/10.4137/GRSB.S9687) PMID: [22654504](https://pubmed.ncbi.nlm.nih.gov/22654504/)
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009). 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25: 2078–9. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
35. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, et al. (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40: 449–454. doi: [10.1038/ng.96](https://doi.org/10.1038/ng.96) PMID: [18344998](https://pubmed.ncbi.nlm.nih.gov/18344998/)
36. Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, et al. (2011). Genome-wide association studies for feedlot and growth traits in cattle. *J Anim Sci* 89: 1684–1697. doi: [10.2527/jas.2010-3079](https://doi.org/10.2527/jas.2010-3079) PMID: [21239664](https://pubmed.ncbi.nlm.nih.gov/21239664/)
37. Henshall JM, Hawken RJ, Dominik S, Barendse W. (2012). Estimating the effect of SNP genotype on quantitative traits from pooled DNA samples. *Genet Sel Evol* 44: 12. doi: [10.1186/1297-9686-44-12](https://doi.org/10.1186/1297-9686-44-12) PMID: [22507187](https://pubmed.ncbi.nlm.nih.gov/22507187/)
38. Tan AC, Fan JB, Karikari C, Bibikova M, Garcia EW, et al. (2008). Allele-specific expression in the germline of patients with familial pancreatic cancer: an unbiased approach to cancer gene discovery. *Cancer Biol Ther* 7: 135–44. PMID: [18059179](https://pubmed.ncbi.nlm.nih.gov/18059179/)
39. Chomczynski P, Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162: 156–9. PMID: [2440339](https://pubmed.ncbi.nlm.nih.gov/2440339/)
40. Martin MC. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*, North America. 17. (<http://journal.embnet.org/index.php/embnetjournal/article/view/200/479>).
41. Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–60. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
42. Anders S, Pyl PT, Huber W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 1: 166–9.
43. Hu Z, Park CA, Reecy JM. (2016). Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Research*, 44: D827–D833. doi: [10.1093/nar/gkv1233](https://doi.org/10.1093/nar/gkv1233) PMID: [26602686](https://pubmed.ncbi.nlm.nih.gov/26602686/)
44. KBioscience (2009). PrimerPicker Lite for KASPar v.0.26. Hoddesdon: KBioscience Ltd.